The diagnostic performance of scoring systems to identify symptomatic colorectal cancer compared to current referral guidance

Tom Marshall[§1] Robert Lancashire[1] Debbie Sharp[2] Tim J Peters[2] KK Cheng[1] and William Hamilton[2]

1. Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham, B15 2TT.
2. Academic Unit of Primary Health Care, Department of Community Based Medicine, University of Bristol, 25-27 Belgrave Road, Bristol, BS8 2AA, UK.

Tom Marshall[§], senior lecturer. E-mail: T.P.Marshall@bham.ac.uk

Robert Lancashire, data scientist. E-mail: R.J.Lancashire@bham.ac.uk

Debbie Sharp, professor. E-mail: Debbie.Sharp@bristol.ac.uk

Tim J Peters, professor. E-mail: Tim.Peters@bristol.ac.uk

KK Cheng, professor. E-mail: K.K.Cheng@bham.ac.uk

William Hamilton, general practitioner and senior lecturer. E-mail: w.hamilton@bristol.ac.uk

[§]Corresponding author

Keywords:
Colorectal cancer, Diagnostic characteristics, Symptoms, Primary health care

Word count 3655; Abstract 239; Tables 5; Figures 2; References 19

Abstract

Objectives

To determine the discrimination characteristics of a new algorithm and two existing symptom scoring systems for identification of patients with suspected colorectal cancer.

Design

Derivation of algorithm by a case-control study and assessment of discrimination characteristics using receiver operating characteristics (ROC) curves. Three colorectal cancer scoring systems were investigated. The Bristol-Birmingham (BB) equation, which we derived from a large primary care dataset; the CAPER score, previously derived from a primary care case-control study and a symptom score derived from National Institute of Clinical Excellence (NICE) guidance for urgent referral of symptomatic patients. Their discrimination characteristics were investigated in two datasets: the BB derivation dataset and the CAPER score derivation dataset. The main analyses were ROC curves and the areas under them for all three algorithms in both datasets.

Setting

Electronic primary care databases

Main outcome measures

Diagnosis of colorectal cancer

Results

In the BB dataset, areas under the curve were: BB equation 0.83 (95% CI: 0.82 to 0.84); CAPER 0.79 (95% CI: 0.79 to 0.80); the NICE guidelines 0.65 (95% CI: 0.64 to 0.66). In the CAPER dataset, areas under the curve were: BB 0.92 (95% CI 0.91 to 0.94); CAPER 0.91

(95% CI 0.89 to 0.93); NICE guidelines 0.75 (95% CI 0.72 to 0.79). In subjects under 50 the discrimination characteristics of NICE referral guidelines were no better than chance.

Conclusions

Both multivariable symptom scoring systems performed significantly better than NICE referral guidelines.

What this paper adds

What is already known on this subject

Despite the availability of screening, the majority of colorectal cancers will continue to be diagnosed after presentation with symptoms

Selection of patients for further investigation depends on combining information from a number of symptoms and signs

The existing symptom scoring systems (NICE guidance and the CAPER score) to help primary care physicians identify which patients should be referred for further investigation have not been evaluated.

What this study adds

Both the new Bristol-Birmingham equation and the CAPER score are markedly better at discriminating between patients with and without colorectal cancer than current NICE guidelines

In patients aged under 50 years current NICE guidelines for urgent referral have no ability to discriminate between patients with and without colorectal cancer

Introduction

Colorectal cancer remains an important cause of death in the UK. Poorer survival rates in international comparisons may be influenced by later presentation.[1-3] Delays in presentation to medical care and diagnosis are well recognised.[4] Despite introduction of a national screening programme in the UK for those aged 60 to 69, the majority of cancers will continue to be diagnosed clinically because most cancers occur after this age, some decline screening, and screening does not detect all cancers.[5]

Diagnosis of colorectal cancer is difficult because the condition is relatively uncommon in primary care and the symptoms are also features of more common, benign conditions. A typical full-time general practitioner will diagnose only one new case annually.[6] Colonoscopy is the main diagnostic test for suspected colorectal cancer, but this is an uncomfortable procedure, requires referral and has a small rate of important complications.

There are a number of different approaches to helping general practitioners select patients for further investigation. Single symptoms have a low specificity for colorectal cancer, but symptom pairs may have more useful test characteristics.[7-8] The National Institute for Clinical Excellence (NICE) published national Referral Guidelines for Suspected Cancer in 2000, and updated these in 2005.[9] These use an algorithm based on age and the presence of certain clinical features. However the guidelines concentrate on typical presentations of cancer; it has been argued that they may even delay diagnosis in patients with atypical presentations.[10] Although the guidelines do not recommend referral of patients with constipation or abdominal pain, these features are clearly associated with cancer.[5 10] It is possible that current referral guidance will reinforce the finding that diagnostic delay is most common in patients who present with change of bowel habit.[11]

The CAPER score is a risk scoring system using multiple presenting symptoms.[12] It was derived from a primary care based case-control study in a single primary care trust.[13-14] The cases were 349 colorectal cancers diagnosed in persons aged over 40 in Exeter between 1998

and 2002. Five age, sex and practice matched controls were obtained for each case. In the CAPER scoring system, some clinical features – abnormal rectal examination, severe anaemia or rectal bleeding – are on their own considered sufficiently high risk to warrant investigation. The CAPER score itself is intended for use with patients without these high-risk features, but who have multiple low-risk symptoms. The score seeks to identify those at higher risk from this low-risk pool. The weaknesses of the CAPER study were that it was undertaken in a single geographical area, used paper-based records and was relatively small.[12]

This paper describes the derivation of the Bristol-Birmingham (BB) equation and compares its performance to the NICE referral guidelines and the CAPER score, using the two different datasets used to derive the BB equation and the CAPER score.

Materials and Methods

Derivation of the BB equation: identification of cases, controls and variables

The BB equation was derived from data provided by The Health Improvement Network (THIN), a national database of electronic primary care records.[15] THIN includes all consultations, prescriptions, diagnoses and primary care investigations for all patients in participating practices. There are 2.2 million currently active patients in over 300 practices, distributed across all regions of the U.K.

We have previously reported the derivation of risk estimates for colorectal cancer based on symptom pairs.[16] The multivariable model described here is an extension of this. Cases were all patients aged 30 years or older with a diagnosis of colorectal cancer between January 2001 and July 2006 (data before this was excluded as direct laboratory transfer of haemoglobin values began around 2000). Seven controls per case, matched for practice, sex and age were selected using a computerised random number sequence (seven being the standard number offered in THIN database studies). Where possible, controls were matched to the same age in years as cases (this was possible for 96.4%): the remainder were matched within one year, two years etc., up to a maximum of five years. Only cases and controls with at least two years of electronic records before the date of diagnosis of the case (the index date) were used. For a small number of very old cases in small practices, fewer than seven controls could be found. THIN staff performed these stages.

Read codes for 24 clinical features of colorectal cancer were selected (list available from authors) and identified in the medical records. Only the two years before the index date were studied. A new prescription may be a proxy for a symptom, such as a laxative for constipation. Initially, symptoms and prescriptions for such symptoms were studied separately. Weight loss was calculated from the most recent and previous weights and divided into two categories: ≥10% weight loss and 5% to 10% weight loss. Most patients did not have two recorded weights, so were labelled unknown. There is a specific Read code for weight

loss – doctors had at times used it without recording an actual weight; it was studied separately. Anaemia was defined if the most recent record of haemoglobin was less than 11g/dl in women and less than 12g/dl in men.

Three risk markers (as opposed to diagnostic features) were also studied: diabetes, obesity and deprivation. For diabetes, patients were considered to be exposed if they had been diagnosed with diabetes at any time before the index date. Obesity was defined as a body mass index greater than 30kg/m^2 within two years of the index date. Each patient was allocated a deprivation quintile based on the Townsend score of their postcode. Irritable bowel syndrome is a potential misdiagnosis: we identified all patients with a record of this diagnosis at any time.

Derivation of the BB equation: data analysis

The initial analysis used univariable conditional logistic regression. Several variables were combined after initial analyses. As the odds ratios for diarrhoea, constipation and abdominal pain were similar to the odds ratios for the related prescription for these symptoms, the pairs of variables were combined. From now on when we use the terms diarrhoea, constipation or abdominal pain, we are referring to either the symptom or a new related prescription for the symptom. The odds ratios for weight loss without a recorded weight were similar to that for ≥10% weight loss; for a haemoglobin result >14g/dl similar to that for no haemoglobin result; for a MCV >85 fl similar to that for no MCV result. These categories were combined.

Variables associated with colorectal cancer with a p-value ≤0.1 were entered into multivariable conditional logistic regression analyses. The model included the following variables: constipation, diarrhoea, change in bowel habit, flatulence, a diagnosis of irritable bowel syndrome, abdominal pain, rectal bleeding, haemoglobin concentration (in 1g/dl bands), microcytosis in two bands (mean cell volume <80 fl and 80-84.99 fl), weight loss (<5%, 5% to 9.9% and ≥10%), venous thrombosis or thromboembolism, diabetes and obesity. The first stage of the multivariable analysis included only symptoms in clinically related

groups: intestinal motility symptoms included constipation, diarrhoea, change in bowel habit and flatulence; pain symptoms included irritable bowel syndrome and abdominal pain; bleeding symptoms included rectal bleeding, anaemia and microcytosis; systemic symptoms included weight loss and thromboembolism; obesity symptoms included diabetes and obesity. The next stage included included all symptoms. Variables where the p-value was greater than 0.05 at any stage, including the final model, were excluded, though these were checked by adding them individually to the final model, using likelihood ratio testing.

Discrimination characteristics

To investigate the discrimination characteristics of the BB equation, individuals in the test dataset were allocated a score equal to their multivariable odds ratio. Because patients with a positive faecal occult blood test, or an abnormal rectal examination, or an abdominal mass unquestionably qualify for further investigation they were allocated an arbitrary maximum score of 100 points in order to make them the highest priority for referral. In this way the score reflected relative priority for further investigation.

The CAPER scores of participants in this study were derived from the presence or absence of six features of colorectal cancer – constipation (25 points), diarrhoea (10), loss of weight (20), abdominal pain or tenderness (15), and one laboratory finding - low haemoglobin (10-11.9g/dl 30 points; 12-12.9g/dl 20 points). The CAPER system was derived for use in a population with only low-risk symptoms, with investigation suggested for scores of $\geq$35 points. Patients with abnormal rectal examination, severe anaemia (haemoglobin <10 g/dl) or rectal bleeding were considered to need referral, so were also allocated an arbitrary score of 100 (Table 2).

The NICE guidelines offer a binary choice: urgent referral or no urgent referral – on the basis of a series of categorical variables: age over 40 years, age over 60 years, sex, menopausal status, diarrhoea (looser stools or increasing stool frequency) of six weeks' duration, rectal bleeding, abdominal mass, abnormal rectal examination and anaemia. Urgent referral is

recommended for patients aged over 60 years with increased stool frequency or with rectal bleeding for six weeks; aged over 40 years with increased stool frequency and rectal bleeding for six weeks; with an abdominal mass or abnormal rectal examination; iron deficiency anaemia (Hb <11mg/dl) in men; iron deficiency anaemia (Hb <10mg/dl) in postmenopausal women.

In the main analysis (NICE 3) we interpreted NICE guidance as follows. The occurrence of two consultations with diarrhoea or change in bowel habit separated by more than 35 but less than 120 days was considered to indicate increased stool frequency for at least six weeks. Consultations separated by more than 120 days are likely to be separate episodes. A single consultation with rectal bleeding was taken to indicate rectal bleeding for at least six weeks. We again assigned a score of 100 points for abdominal mass, positive faecal occult blood or abnormal rectal examination for consistency. We assigned one point for the following features in which urgent investigation is advised: diarrhoea plus rectal bleeding and aged over 40 years; rectal bleeding and aged over 60; diarrhoea and aged over 60; haemoglobin <11g/dl with microcytosis in a man; and haemoglobin <10g/dl with microcytosis in a postmenopausal woman (age>52 was taken as a proxy for being postmenopausal). The score thus rose with the number of qualifying symptoms (Table 2). In the CAPER dataset, the mean cell volume was not available, so the requirement for microcytosis was dropped; therefore a haemoglobin <10 g/dl in a man was allocated one point.

Two sensitivity analyses used different interpretations of the NICE guidelines. In NICE 1, a single consultation with diarrhoea or change in bowel habit was taken to indicate diarrhoea for six weeks. In NICE 2, two consultations separated by more than 35 but less than 120 days but not change in bowel habit, were taken to indicate diarrhoea for six weeks.

The receiver operating characteristics of a prediction model are usually superior in the dataset from which it was derived. To avoid this "home advantage", two datasets were used to assess the predictive power of the equations: the dataset used to derive the BB equation (described above) and the CAPER dataset. The CAPER dataset includes 349 colorectal cancer cases and

1744 age and sex matched controls.[8] The mean age of cases was 71.9 years (range 40 to 96) and 50.1% of the dataset was male. The CAPER dataset was obtained by searching both paper and electronic primary care records for symptoms. In the CAPER dataset, weight loss was recorded as only present or absent; therefore this was taken to be equivalent to a >10% weight loss.

Receiver operating characteristics (ROC) curves were constructed and the area under the curve was determined for the three prediction models in both datasets. The large size of the THIN dataset allowed us to undertake extensive sensitivity analyses. We repeated ROC curves for men and women, for each year of diagnosis of cancer from 2001 to 2007 and in ten year age bands. To determine whether allocating a "mandatory referral" score to abdominal masses, positive faecal occult bloods or abnormal rectal examinations affected the findings, we excluded cases and controls with these features and repeated the analysis. Because the CAPER score was derived from persons aged 40 and over we also repeated the analysis in persons aged over 40.

Yield

We estimated the yield of colorectal cancers using these systems by calculating the positive predictive values (PPVs) at selected points of the ROC curves, using Bayes' theorem (posterior odds = prior odds×likelihood ratio).[17] To compare the systems, we used the points on the three ROC curves with the same sensitivity. We derived the prior odds from national incidence rates for 2006, stratified by age and sex.[18]

Results

We identified 5,477 cases and 38,314 controls in a total of 317 practices. Their mean age was 70.6 years (range 30 to 105) and 53.1% were male. Demographic details of subjects are shown in Table 1.

Derivation of the BB equation

In the univariable analyses, positive faecal occult bloods (odds ratio 24.5, 95% CI 5.1 to 118), abnormal rectal examination (101, 13.3 to 765.2) and abdominal mass (35.0, 20.8 to 58.9) were strongly associated with a diagnosis of colorectal cancer. (Table 3) As these features warrant investigation *per se* they were dropped from further modelling. A family history of colorectal cancer was recorded in only seven cases and eight controls (odds ratio 6.13 95% CI 2.22 to 16.9).

Multivariable analysis was carried out using 13 variables, plus the deprivation quintile. In multivariable analyses none of flatulence, irritable bowel syndrome, diabetes, obesity, thromboembolism, or deprivation quintile was independently associated with cancer. The final model therefore included eight variables: constipation, diarrhoea, change in bowel habit, abdominal pain, haemoglobin concentration mean cell volume and weight loss (Table 4).

Discrimination characteristics 1. In the BB dataset

Table 2 summarises the way in which scores were derived from the three equations. In the THIN dataset the area under the curve for the BB equation was 0.83 (95% CI: 0.82 to 0.84) and for CAPER was 0.79 (95% CI: 0.79 to 0.80). The area under the curve for the NICE 3 interpretation of the NICE guidelines was 0.65 (95% CI: 0.64 to 0.66) and was consistently superior to NICE 1 and NICE 2. (Figure 1) Excluding patients under 40 made little difference, and excluding those with abdominal mass, positive faecal occult blood and abnormal rectal examination only made modest differences. In all analyses, the BB equation and CAPER score remained superior to NICE.

For the BB equation and NICE guidelines the areas under the curve were similar in men and women. The CAPER score performed slightly better in women than men. The BB and CAPER scores performed similarly at all ages. However no interpretation of the NICE

guidelines performed better than chance at ages under 50. NICE guidelines performed best at age 80 to 89 years.

Discrimination characteristics 2. In the CAPER dataset

In the CAPER dataset, the areas under the curve were BB 0.92 (95% CI 0.91 to 0.94); CAPER 0.91 (95% CI 0.89 to 0.93); NICE 1 0.76 (95% CI 0.73 to 0.80). (Table 5) In this dataset the area under the curve for NICE 1 was greater than for NICE 3. (Figure 2)

Yield

Using the more conservative estimate of discrimination characteristics derived from the THIN dataset, NICE 3 has a sensitivity of 0.327 and a specificity of 0.974, giving a positive likelihood ratio of 12.5 and a positive predictive value (yield) of 3.1% at age 70-74. A point on the CAPER and BB ROC curves for the same age was selected to have the same sensitivity. These points had the following characteristics: CAPER – positive likelihood ratio 13.4 and PPV 3.3%.; BB – positive likelihood ratio 14.7 and PPV 3.7%. (Table 6)

Discussion

In both datasets the overall discrimination characteristics of the BB equation were consistently slightly better than those of the CAPER score and both were superior to any of the interpretations of current guidance. NICE guidelines performed no better than chance in subjects aged under 50.

Weakness and strengths

The performance of all methods of identifying colorectal cancer was better in the CAPER dataset than in the THIN dataset. The CAPER dataset has some probable advantages: cases were identified from the cancer registry and clinical features of colorectal cancer were identified from both paper and electronic records.[12] The overall standardised incidence of cancer in THIN is consistent with cancer registry data, but there is some under-recording, probably as a result of misclassification of solid tumours. Standardised incidence ratios for colorectal cancer range from 0.69 to 0.84 between 2000 and 2006.[19] Data linkage to cancer registry records could improve the recording of outcomes.

We did not use the 'free text' comments in the THIN records.[20] This will have meant some symptom recording was missed, though there is no reason to believe this would be more common in cases than controls.[12]

Comparison with previous literature

NICE guidelines may perform less well than the BB equation and CAPER because they include fewer predictor variables, some of which only apply at certain ages. This means that NICE guidelines perform well for the minority of colorectal cancers with typical features (the first part of the ROC curve), but less well for the majority of cases with low risk but not no-risk symptoms.[21] Variables absent from NICE guidelines include constipation, loss of weight

and abdominal pain. Diarrhoea, rectal bleeding and anaemia are part of CAPER and BB at all ages, though have an age-restriction within NICE. The BB equation includes two further variables, microcytosis and change in bowel habit and divides haemoglobin level and weight loss into subcategories. Change in bowel habit is an important predictor of colorectal cancer, and clearly doctors use this term differently to either diarrhoea or constipation. One criticism of NICE guidance is that only the minority of patients with colorectal cancer have a high risk symptom before diagnosis, with the majority experiencing constipation, diarrhoea or abdominal pain (or a combination of these).[5] Thus, it is not surprising that NICE fails to identify such patients, and that survival from colorectal cancer has improved little since they were published.

Other referral guidelines and symptom scoring systems might be investigated in a similar way. For example, the Selva score was derived from patients referred to secondary care for investigation and makes use of a consultation questionnaire to elicit symptoms.[22] It has been reported to have an area under the curve of 0.76 in a population of patients referred for investigation of suspected colorectal cancer[23] though it had a very poor performance in the one reported primary care study.[24]

Primary care datasets are an invaluable resource for investigating the discrimination characteristics of referral guidelines and therefore of informing recommendations. Multivariable models to guide referral have much better discrimination characteristics than current NICE guidelines and so have the potential to significantly improve the selection of patients for further investigation of colorectal cancer symptoms. There is a strong case for a cohort analysis to derive a statistical model of cumulative incidence of colorectal cancer.

Acknowledgements

We wish to thank THIN staff, especially Mary Thompson for their help and diligence.

Competing interests

All authors declare that the answer to the questions on your competing interest form are all No and therefore have nothing to declare

Funding and sponsorship

Project funding from Cancer Research UK, and sponsorship by the University of Bristol. CRUK reference number: C12345/A7502. Neither body had a role in study design, data collection, analysis or writing of the report. The researchers are independent of the funding body.

Contributors

TM and WH drafted the original study protocol in collaboration with DS, TP and KK. RL undertook the analysis. The remaining authors assisted in design, funding and supervision of the study. All authors contributed to revising the article, which was drafted by TM, who is the guarantor.

Ethical approval

London MREC; 06/MRE02/75

References

1. Thomson CS, Forman D. Cancer survival in England and the influence of early diagnosis: what can we learn from recent EUROCARE results[quest]. *Br J Cancer* 2009;101(S2):S102-S09.
2. Moller H, Linklater KM, Robinson D. A visual summary of the EUROCARE-4 results: a UK perspective. *Br J Cancer* 2009;101(S2):S110-S14.
3. Richards MA. The size of the prize for earlier diagnosis of cancer in England. *Br J Cancer* 2009;101(S2):S125-S29.
4. Thorne K, Hutchings H, Elwyn G. The effects of the Two-Week Rule on NHS colorectal cancer diagnostic services: A systematic literature review. *BMC Health Services Research* 2006;6:43.
5. Hamilton W. Five misconceptions in cancer diagnosis. *British Journal General Practice* 2009;59:441-47.
6. Summerton N. *Diagnosing cancer in primary care*. Abingdon: Radcliffe Medical Press, 1999.
7. Bjerregaard NC, Tottrup A, Sorensen HT, Laurberg S. Diagnostic value of self-reported symptoms in Danish outpatients referred with symptoms consistent with colorectal cancer. *Colorectal Disease* 2007;9:443-51.
8. Hamilton W, Round A, Sharp D, Peters T. Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *British Journal of Cancer* 2005;93:399-405.
9. NICE. *Referral Guidelines for suspected cancer*. London: NICE, 2005.
10. Jones R, Rubin G, Hungin P. Is the two week rule for cancer referrals working? *BMJ* 2001;322:1555-56.
11. Barrett J, Jiwa M, Rose P, Hamilton W. Pathways to the diagnosis of colorectal cancer: an observational study in three UK cities. *Fam. Pract.* 2006;23:15-19.
12. Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* 2009;101(S2):S80-S86.
13. Hamilton W. Derivation of a score for identifying colorectal cancer in primary care. *Gut* 2007;56(Suppl 2):A49.
14. Khan N, NCRI Colorectal Clinical Studies Group. Implementation of a diagnostic tool for symptomatic colorectal cancer in primary care: a feasibility study. doi:10.1017/S1463423608000996. *Primary Health Care Research & Development* 2009;10:54-64.
15. Andrew Maguire, Betina T. Blak, Mary Thompson. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiology and Drug Safety* 2009;18(1):76-83.
16. Hamilton W, Lancashire R, Sharp D, Peters T, Cheng K, Marshall T. The risk of colorectal cancer with symptoms at different ages and between the sexes: a case-control study. *BMC Medicine* 2009;7:17.
17. Knottnerus JA. *The evidence base of clinical diagnosis*. London: BMJ Books, 2002.
18. Northern and Yorkshire Cancer Registry and Information Service. 2006.
19. Haynes K, Forde K, Schinnar R, Wong P, Strom B, Lewis J. Cancer incidence in The Health Improvement Network. *Pharmacoepidemiology and Drug Safety* 2009;18(8):730-36.
20. Tate A, Martin A, Murray-Thomas T, Anderson S, Cassell J. Determining the date of diagnosis - is it a simple matter? The impact of different approaches to dating

diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Medical Research Methodology* 2009;9:42.
21. Hamilton W. Cancer diagnosis in primary care. *British Journal of General Practice* 2010;60:121-28.
22. Selvachandran S, Hodder R, Ballal M, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *Lancet* 2002;360:278-83.
23. Ballal MS, Selvachandran SN, Maw A. Use of a patient consultation questionnaire and weighted numerical scoring system for the prediction of colorectal cancer and other colorectal pathology in symptomatic patients: A prospective cohort validation study of a Welsh population. *Colorectal Disease* 2010;12:407-14.
24. Soin G, Armitage J, McKay J, Ballal M, Selvachandran S, Cade D. Colorectal symptoms in the community - a ticking time bomb? *Gut* 2004;53(suppl_3):Abstract 069.

Table 1: Demographic details of cases and controls in the Bristol-Birmingham database

Age band	Number of cases						Number of controls					
	Male (%)		Female (%)		Total (%)		Male (%)		Female (%)		Total (%)	
30-39	21	(0.4%)	29	(0.5%)	50	(0.9%)	147	(0.4%)	203	(0.5%)	350	(0.9%)
40-49	104	(1.9%)	109	(2.0%)	213	(3.9%)	728	(1.9%)	763	(2.0%)	1,491	(3.9%)
50-59	425	(7.8%)	317	(5.8%)	742	(13.5%)	2,975	(7.8%)	2,219	(5.8%)	5,194	(13.6%)
60-69	761	(13.9%)	572	(10.4%)	1,333	(24.3%)	5,334	(13.9%)	4,002	(10.4%)	9,336	(24.4%)
70-79	1,021	(18.6%)	787	(14.4%)	1,808	(33.0%)	7,160	(18.7%)	5,509	(14.4%)	12,669	(33.1%)
80-89	519	(9.5%)	634	(11.6%)	1,153	(21.1%)	3,667	(9.6%)	4,449	(11.6%)	8,116	(21.2%)
90+	60	(1.1%)	118	(2.2%)	178	(3.2%)	350	(0.9%)	808	(2.1%)	1,158	(3.0%)
Total	2,911	(53.1%)	2,566	(46.9%)	5,477	(100.0%)	20,361	(53.1%)	17,953	(46.9%)	38,314	(100.0%)

Source: The Bristol-Birmingham database was extracted from the THIN database of electronic primary care records

Table 2: Coefficients for the BB equation, scores for CAPER and for three interpretations of NICE guidelines on urgent referral of patients with suspected colorectal cancer

Predictor variable	BB	CAPER	NICE 1	NICE 2	NICE 3
Constipation	2.06	25	Not applicable	Not applicable	Not applicable
Diarrhoea	2.38	10 per episode maximum of 4 episodes	1 if aged >40 with rectal bleeding. 1 if aged >60.	1 if two consultations 35 to 119 days apart and aged >60. 1 if two consultations 35 to 119 days apart & aged >40 and rectal bleeding.	1 if two consultations 35 to 119 days apart and aged >60. 1 if two consultations 35 to 119 days apart & aged >40 and rectal bleeding.
Change in bowel habit	13.83	Not applicable	Not applicable	Not applicable	
Prescription of antispasmodic or abdominal pain	3.82	15 per episode maximum of 3 episodes	Not applicable	Not applicable	Not applicable
Rectal bleeding	20.11	100	1 if aged >60	1 if aged >60	1 if aged >60
Haemoglobin 13.0 – 13.9 mg/dl	1.33	Not applicable	Not applicable	Not applicable	Not applicable
Haemoglobin 12.0 – 12.9 mg/dl	1.63	20	Not applicable	Not applicable	Not applicable
Haemoglobin 11.0 – 11.9 mg/dl	2.54	30	Not applicable	Not applicable	Not applicable
Haemoglobin 10.0 – 10.9 mg/dl	5.18	30	1 with MCV <80 fl in a man* 1 with MCV<80 fl in a woman >52 years*	1 with MCV <80 fl in a man* 1 with MCV<80 fl in a woman >52 years*	1 with MCV <80 fl in a man* 1 with MCV<80 fl in a woman >52 years*
Haemoglobin 9.0 – 9.9 mg/dl	8.08	100			
Haemoglobin <9.0 mg/dl	15.94	100			
Mean Cell Volume 80 – 84.9 fl	2.71	Not applicable	Not applicable	Not applicable	Not applicable
Mean Cell Volume <80 fl	7.67	Not applicable	See Hb	See Hb	See Hb
Weight loss ≥10% (or recorded)	2.92	20	Not applicable	Not applicable	Not applicable
Weight loss 5% - 9.9%	1.37	Not applicable	Not applicable	Not applicable	Not applicable
Not known	1.21	Not applicable	Not applicable	Not applicable	Not applicable
Faecal Occult Blood	100	100	100	100	100
Abnormal Rectal Examination	100	100	100	100	100
Abdominal Mass	100	100	100	100	100

** MCV criterion not required in CAPER dataset*

Table 3: Initial variables considered for inclusion in the Bristol-Birmingham predictive model and their frequency in cases and controls

Risk	Description	Cases (N = 5,477)		Controls (N = 38,314)		Odds ratio (95% confidence interval)	
		Number (%) with this risk		Number (%) with this risk			
1	New episode of Constipation with or without prescription of a Laxative	684	(12.49%)	1,867	(4.87%)	2.92	(2.66 to 3.22)
2	New prescription of a Laxative (without a record of Constipation)	793	(14.48%)	2,184	(5.70%)	2.92	(2.67 to 3.20)
3	New episode of Diarrhoea with or without prescription of an Antimotility Drug	830	(15.15%)	1,764	(4.60%)	3.82	(3.49 to 4.18)
4	New prescription for an Antimotility Drug (without a record of Diarrhoea)	158	(2.88%)	407	(1.06%)	2.80	(2.32 to 3.38)
5	New episode of Change in Bowel Habit (with Diarrhoea/Antimotility Drug also mentioned)	137	(2.50%)	95	(0.25%)	10.25	(7.87 to 13.33)
6	New episode of Change in Bowel Habit (with Constipation/Laxative also mentioned but no Diarrhoea)	134	(2.45%)	90	(0.23%)	10.85	(8.27 to 14.25)
7	New episode of Change in Bowel Habit (without either Diarrhoea or Constipation mentioned)	344	(6.28%)	190	(0.50%)	13.90	(11.56 to 16.73)
8	New diagnosis of Irritable Bowel Syndrome (with or without prescription of an Antispasmodic)	120	(2.19%)	303	(0.79%)	2.84	(2.29 to 3.52)
9	New prescription for an Antispasmodic (without a record of Irritable Bowel Syndrome or Change in Bowel Habit)	500	(9.13%)	995	(2.60%)	3.78	(3.38 to 4.23)
10	New episode of Rectal Bleeding or Melaena	853	(15.57%)	460	(1.20%)	16.17	(14.26 to 18.33)
11	New episode of Faecal Occult Blood Present without mention of Rectal Bleeding or Melaena	7	(0.13%)	2	(0.01%)	24.50	(5.09 to 117.94)
12	Weight Loss >10% within the past 2 years	183	(3.34%)	356	(0.93%)	3.72	(3.10 to 4.46)
13	Weight Loss 5% to 10% within the past 2 years	245	(4.47%)	924	(2.41%)	1.91	(1.65 to 2.21)
14	Weight Loss recorded within the past 2 years but actual weight not recorded or no weight loss	149	(2.72%)	286	(0.75%)	3.77	(3.08 to 4.61)
15	New episode of Abdominal Pain or Tenderness	1,412	(25.78%)	2,660	(6.94%)	4.84	(4.50 to 5.22)
16	New episode of Abnormal Rectal Examination	15	(0.27%)	2	(0.01%)	101.02	(13.34 to 765.19)
17	New episode of Anaemia with a record of a low Haemoglobin	1,181	(21.56%)	1,030	(2.69%)	11.40	(10.33 to 12.58)
18	New prescription of Iron for without a record of a low Haemoglobin	556	(10.15%)	1,424	(3.72%)	3.03	(2.72 to 3.36)
19	New episode of flatulence	52	(0.95%)	116	(0.30%)	3.17	(2.28 to 4.40)
20	Diabetes (ever diagnosed)	626	(11.43%)	3,679	(9.60%)	1.22	(1.11 to 1.34)
21	Obesity (BMI >30)	600	(10.95%)	3,846	(10.04%)	1.11	(1.01 to 1.22)
22	New diagnosis of Deep Venous Thrombosis or Pulmonary Embolism	24	(0.44%)	74	(0.19%)	2.27	(1.43 to 3.60)
23	New episode of Abdominal mass	86	(1.57%)	19	(0.05%)	34.98	(20.78 to 58.86)
24	Mean Cell Volume <80 fl	761	(13.89%)	284	(0.74%)	23.30	(20.04 to 27.09)

Source: The Bristol-Birmingham database was extracted from the THIN database of electronic primary care records

Table 4: Results of multivariable conditional logistic regression analysis for the Bristol-Birmingham predictive model

Factor	Patients with this factor	Cases		Controls		Univariable analysis		Multivariate all significant factors	
		Number and prevalence (%)		Number and prevalence (%)		Odds Ratio	95% Confidence Interval	Odds Ratio	95% Confidence Interval
	All patients	5,477		38,314					
1	Constipation	1,477	27.0%	4,051	10.6%	3.35	(3.12 to 3.60)	2.06	(1.88 to 2.26)
2	Diarrhoea	988	18.0%	2,171	5.7%	3.75	(3.45 to 4.07)	2.38	(2.14 to 2.66)
3	Change in Bowel Habit	615	11.2%	375	1.0%	13.11	(11.44 to 15.03)	13.83	(11.70 to 16.34)
4	Flatulence	52	0.9%	116	0.3%	3.17	(2.28 to 4.40)		
5	Irritable Bowel Syndrome	135	2.5%	325	0.8%	2.99	(2.44 to 3.67)		
6	Abdominal pain / Antispasmodic	1,629	29.7%	3,121	8.1%	4.94	(4.60 to 5.30)	3.82	(3.49 to 4.18)
7	Rectal bleeding	853	15.6%	460	1.2%	16.17	(14.26 to 18.33)	20.11	(17.35 to 23.32)
8	Haemoglobin ≥14 g/dl or Not Known	2,963	54.1%	30,316	79.1%	1.00		1.00	
	Haemoglobin 13-13.999 g/dl	573	10.5%	3711	9.7%	2.02	(1.82 to 2.24)	1.33	(1.18 to 1.50)
	Haemoglobin 12-12.999 g/dl	517	9.4%	2,484	6.5%	2.98	(2.66 to 3.33)	1.63	(1.42 to 1.87)
	Haemoglobin 11-11.999 g/dl	417	7.6%	1,131	3.0%	5.38	(4.73 to 6.13)	2.54	(2.16 to 2.99)
	Haemoglobin 10-10.999 g/dl	354	6.5%	417	1.1%	12.27	(10.47 to 14.39)	5.18	(4.19 to 6.39)
	Haemoglobin 9-9.999 g/dl	268	4.9%	153	0.4%	23.49	(18.93 to 29.13)	8.08	(6.13 to 10.65)
	Haemoglobin <9	385	7.0%	102	0.3%	50.88	(40.16 to 64.48)	15.94	(11.78 to 21.57)
9	Mean Cell Volume ≥85 fl or Not Known	4,272	78.0%	37,168	97.0%	1.00		1.00	
	Mean Cell Volume 80-84.999 fl	444	8.1%	862	2.2%	4.95	(4.37 to 5.61)	2.71	(2.30 to 3.19)
	Mean Cell Volume <80 fl	761	13.9%	284	0.7%	26.10	(22.42 to 30.39)	7.67	(6.23 to 9.44)
10	No weight loss	752	13.7%	5,588	14.6%	1.00		1.00	
	Weight loss ≥10% or Recorded Weight Loss	351	6.4%	678	1.8%	3.84	(3.30 to 4.46)	2.92	(2.39 to 3.57)
	Weight loss 5% to 10%	210	3.8%	852	2.2%	1.83	(1.55 to 2.17)	1.37	(1.09 to 1.73)
	1 Recording / NK*	4164	76.0%	31,196	81.4%	0.96	(0.88 to 1.05)	1.21	(1.09 to 1.35)
11	DVT or PE	24	0.4%	74	0.2%	2.27	(1.43 to 3.60)		
12	Diabetes	626	11.4%	3,679	9.6%	1.22	(1.11 to 1.34)		
13	Obesity	600	11.0%	3,846	10.0%	1.11	(1.01 to 1.22)		

** Patients with only one or fewer weight in whom it was not possible to calculate a percentage weight loss*

Table 5: Areas under the curve in each age band and both sexes for a diagnosis of colorectal cancer using the BB (Bristol Birmingham) equation, the CAPER score and three algorithms based on NICE guidelines

Tested in the THIN dataset										
Description of test dataset	BB		CAPER		NICE 1		NICE 2		NICE 3	
Both sexes all ages	0.829	(0.822-0.835)	0.793	(0.786-0.800)	0.642	(0.633-0.651)	0.613	(0.604-0.622)	0.651	(0.642-0.659)
Both sexes age ≥40	0.829	(0.823-0.836)	0.793	(0.786-0.801)	0.643	(0.634-0.652)	0.614	(0.605-0.623)	0.652	(0.643-0.661)
Both sexes excluding "must refer"*	0.826	(0.819-0.833)	0.790	(0.782-0.797)	0.635	(0.626-0.644)	0.606	(0.597-0.615)	0.644	(0.635-0.653)
Men	0.824	(0.815-0.833)	0.781	(0.771-0.791)	0.643	(0.631-0.655)	0.610	(0.597-0.622)	0.655	(0.642-0.667)
Women	0.838	(0.829-0.848)	0.812	(0.802-0.822)	0.640	(0.627-0.653)	0.618	(0.604-0.631)	0.646	(0.633-0.659)
30-39 years	0.758	(0.673-0.844)	0.762	(0.676-0.848)	0.520	(0.432-0.608)	0.520	(0.432-0.608)	0.520	(0.432-0.608)
40-49 years	0.839	(0.805-0.873)	0.816	(0.779-0.853)	0.520	(0.477-0.563)	0.511	(0.469-0.553)	0.532	(0.489-0.575)
50-59 years	0.818	(0.798-0.838)	0.796	(0.775-0.817)	0.540	(0.517-0.564)	0.537	(0.514-0.560)	0.567	(0.543-0.590)
60-69 years	0.840	(0.827-0.854)	0.806	(0.791-0.821)	0.659	(0.641-0.677)	0.624	(0.606-0.642)	0.668	(0.650-0.686)
70-79 years	0.828	(0.817-0.840)	0.791	(0.778-0.803)	0.665	(0.649-0.680)	0.631	(0.615-0.646)	0.670	(0.655-0.686)
80-89 years	0.836	(0.822-0.850)	0.792	(0.776-0.808)	0.682	(0.663-0.701)	0.649	(0.630-0.669)	0.686	(0.666-0.705)
90+ years	0.784	(0.746-0.822)	0.751	(0.712-0.791)	0.632	(0.583-0.681)	0.596	(0.547-0.645)	0.626	(0.576-0.675)
Tested in the CAPER dataset										
Description of test dataset	BB		CAPER		NICE 1		NICE 2		NICE 3	
Both sexes all ages	0.922	(0.905-0.938)	0.908	(0.890-0.927)	0.764	(0.732-0.797)	0.717	(0.681-0.752)	0.752	(0.718-0.786)

** Cases and controls with a positive faecal occult blood test, an abdominal mass or an abnormal rectal examination are excluded from the THIN dataset for this analysis.*

Table 6: Annual age specific incidence of colorectal cancer and positive predictive values if a patient meets the referral criteria (a positive test result)

Equation	Sensitivity	Age band	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
		Incidence per 100,000	39.1	69.8	110.6	171.3	247.2	316.0	375.4	369.7
		Likelihood Ratio	Positive predictive value of a positive test result							
NICE 1	34.1%	5.7	0.2%	0.4%	0.6%	1.0%	1.4%	1.8%	2.1%	2.1%
NICE 2	24.5%	13.3	0.5%	0.9%	1.5%	2.3%	3.3%	4.2%	5.0%	4.9%
NICE 3	32.7%	12.5	0.5%	0.9%	1.4%	2.2%	3.1%	4.0%	4.7%	4.7%
CAPER	32.6%	13.4	0.5%	0.9%	1.5%	2.3%	3.3%	4.2%	5.0%	5.0%
BB	33.5%	14.7	0.6%	1.0%	1.6%	2.5%	3.7%	4.7%	5.6%	5.5%

Source: Annual incidence for England in 2006 obtained from Northern and Yorkshire Cancer Registry and Information Service

Figure 1: Receiver operating characteristics curves for a diagnosis of colorectal cancer in the THIN dataset using the Bristol Birmingham equation, the CAPER score and three algorithms based on NICE guidelines

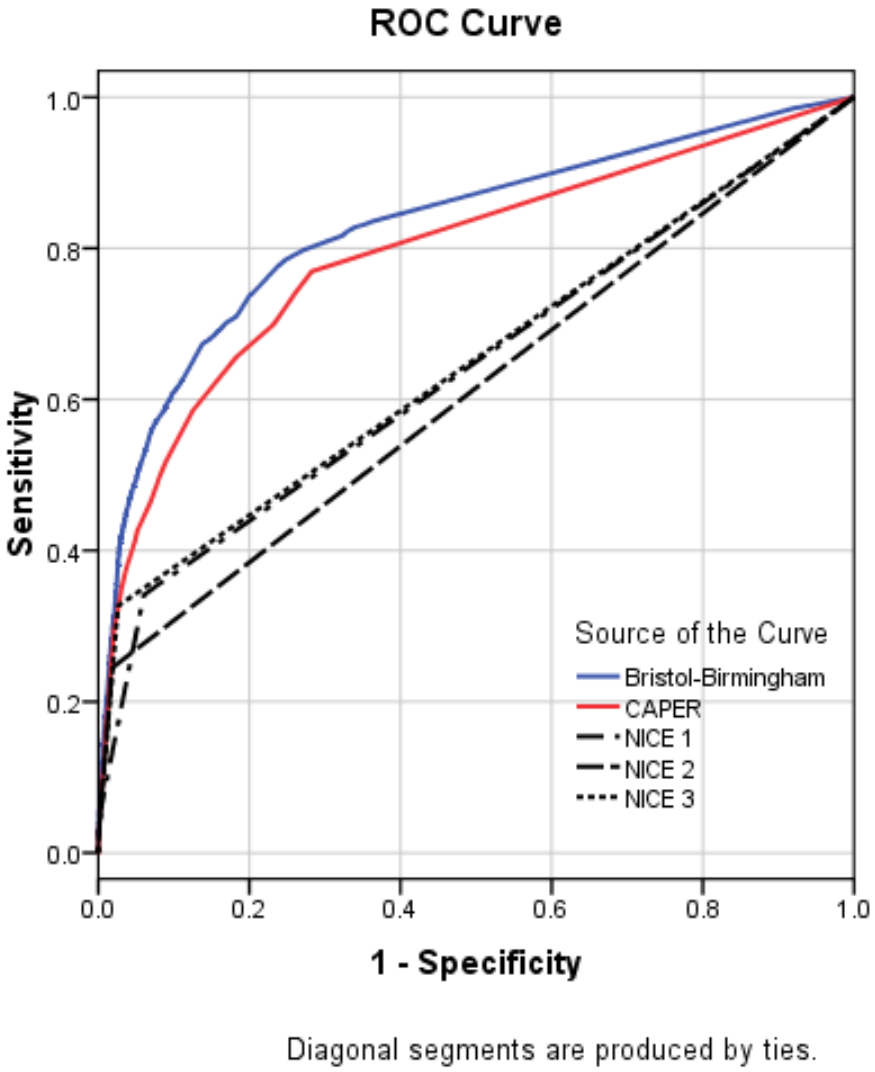
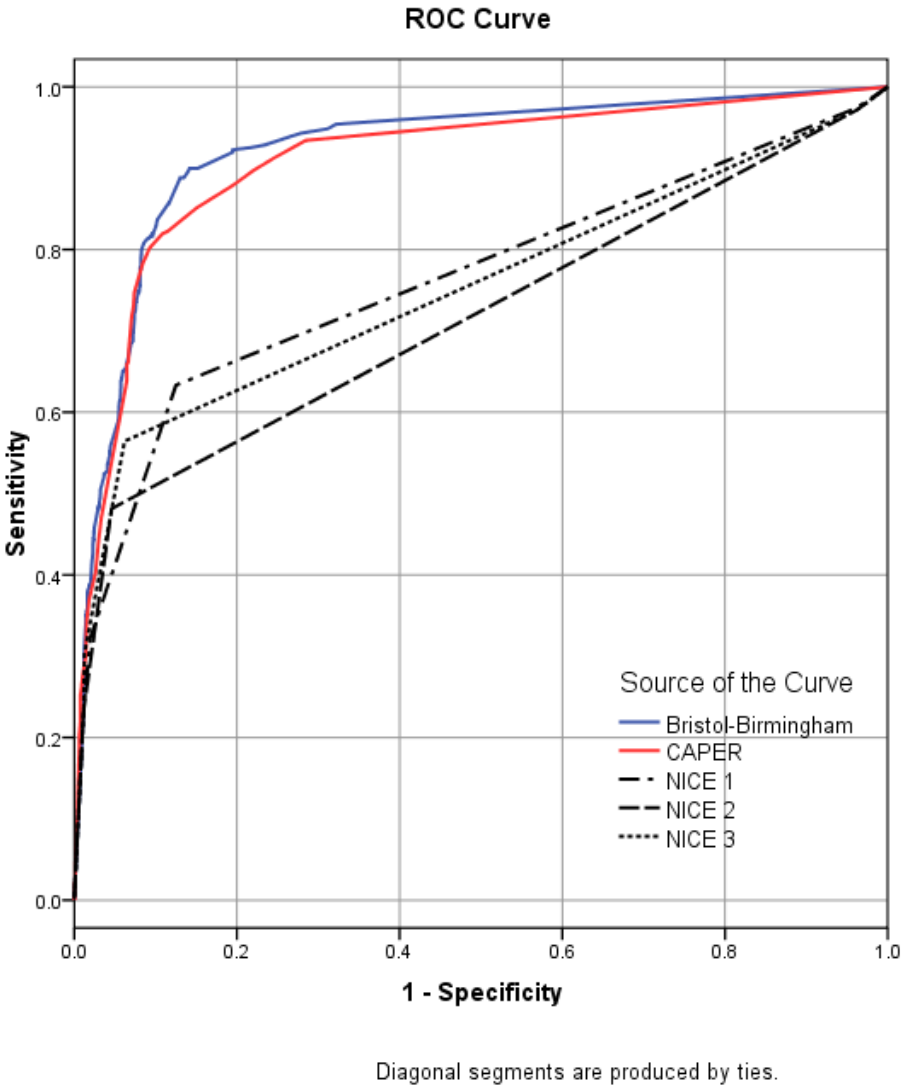


Figure 2: Receiver operating characteristics curves for a diagnosis of colorectal cancer in the CAPER dataset using the Bristol Birmingham equation, the CAPER score and three algorithms based on NICE guidelines



Diagonal segments are produced by ties.